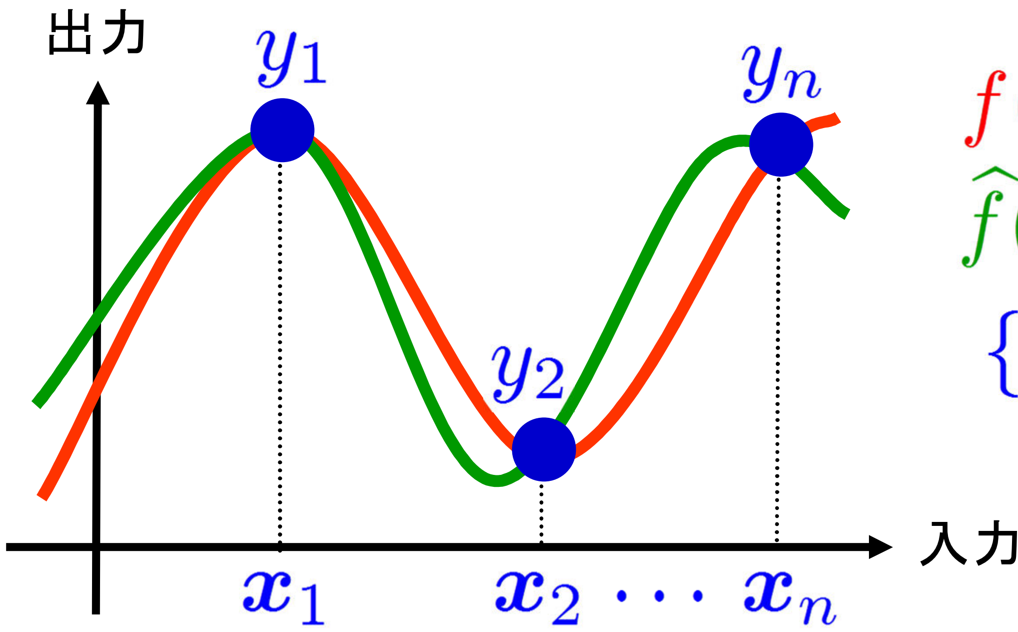
ロバスト回帰
（６章）

杉山将・本多淳也
sugi@k.u-tokyo.ac.jp, jhonda@k.u-tokyo.ac.jp
http://www.ms.k.u-tokyo.ac.jp

解答付きver

回帰 = 関数近似



$f(\boldsymbol{x})$: 学習したい真の関数

$\widehat{f}(\boldsymbol{x})$: 学習結果の関数

$\{(\boldsymbol{x}_i, y_i)\}_{i=1}^n$: 訓練標本

$$y_i = f(\boldsymbol{x}_i)\ (+\mathrm{noise})$$

訓練標本から真の関数にできるだけ近い関数を求める

パラメータに関する線形モデル

- **線形モデル**:

$$f_{\boldsymbol{\theta}}(\boldsymbol{x}) = \sum_{j=1}^{b} \theta_j \phi_j(\boldsymbol{x})$$

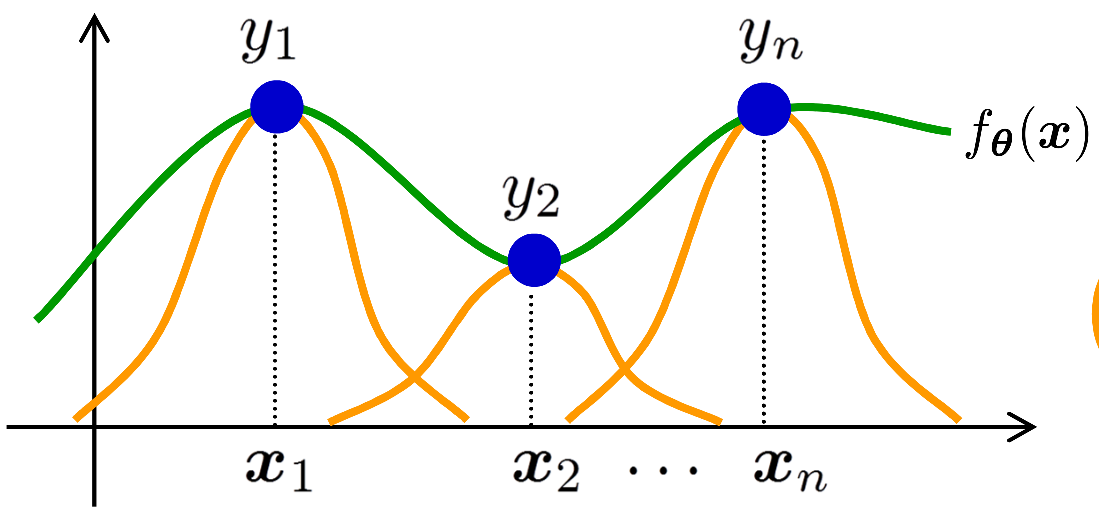
$\{\phi_j(\boldsymbol{x})\}_{j=1}^{b}$:基底関数

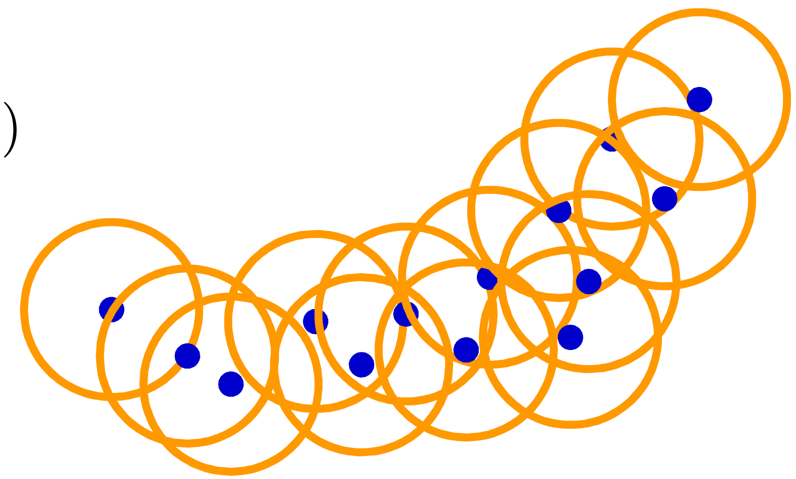
- **カーネルモデル**:

$$f_{\boldsymbol{\theta}}(\boldsymbol{x}) = \sum_{j=1}^{n} \theta_j K(\boldsymbol{x}, \boldsymbol{x}_j)$$

ガウスカーネル

$$K(\boldsymbol{x}, \boldsymbol{c}) = \exp\left(-\frac{\|\boldsymbol{x} - \boldsymbol{c}\|^2}{2h^2}\right)$$

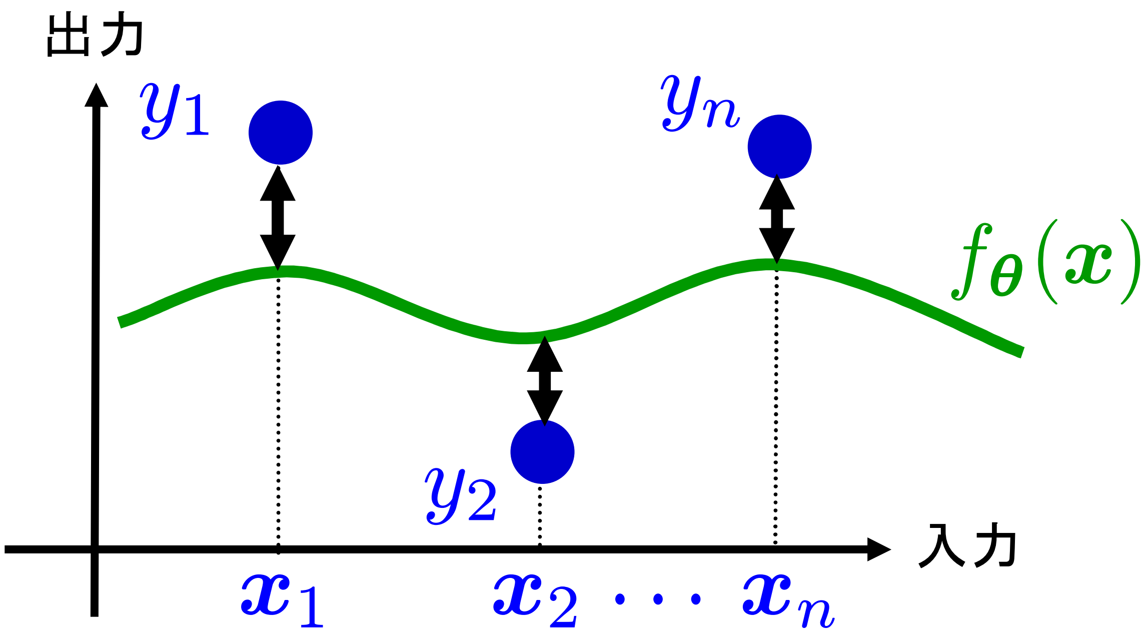




最小二乗回帰

- 訓練出力との二乗誤差を最小にする:

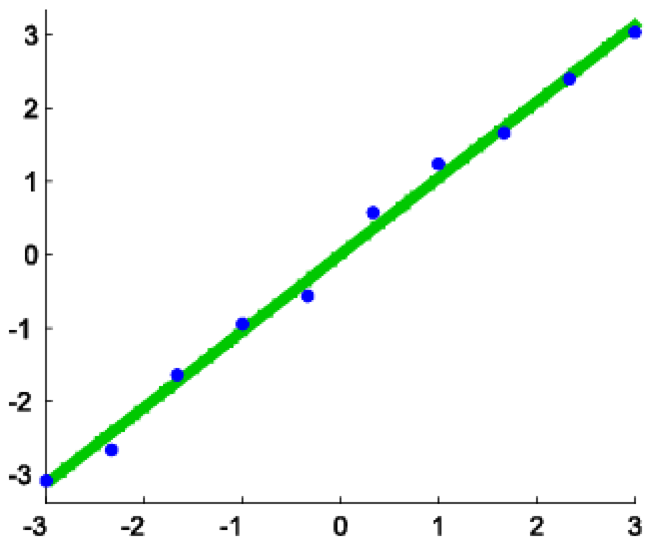
$$\min_{\boldsymbol{\theta}} \sum_{i=1}^{n} \left(f_{\boldsymbol{\theta}}(\boldsymbol{x}_i) - y_i \right)^2$$



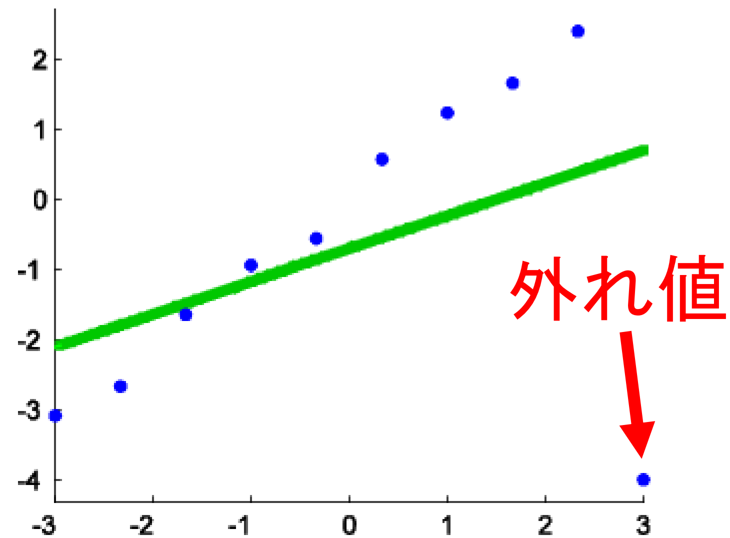
最小二乗回帰の問題点

- たった一つの外れ値が，学習結果を大きく変えてしまう！

$$f_{\boldsymbol{\theta}}(x) = \theta_1 + \theta_2 x$$



最小二乗回帰
（外れ値なし）

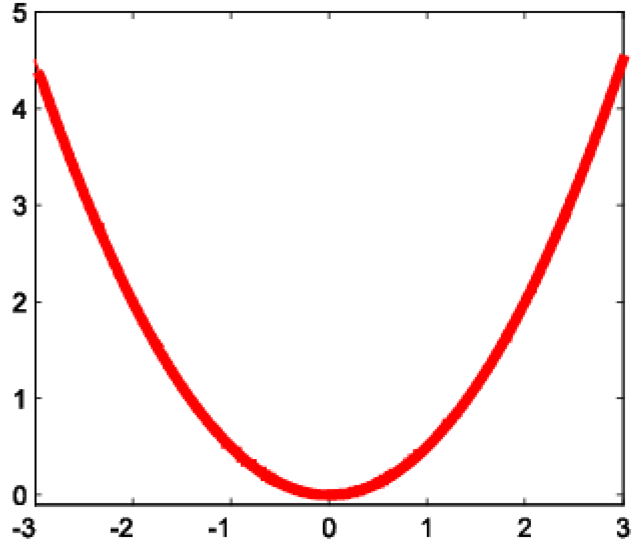


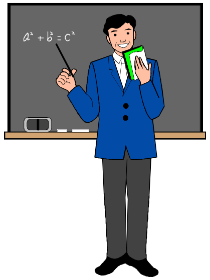
最小二乗回帰
（外れ値あり）

ℓ_2-損失関数

$$\sum_{i=1}^{n}\left(f_{\boldsymbol{\theta}}(\boldsymbol{x}_i)-y_i\right)^2$$

- 最小二乗回帰: 訓練出力との適合のよさを ℓ_2-損失関数で測る
- 外れ値は“二乗の強さ”で学習結果に影響を及ぼす
- 学習結果は安定させるには外れ値の影響を小さくする必要がある





講義の流れ

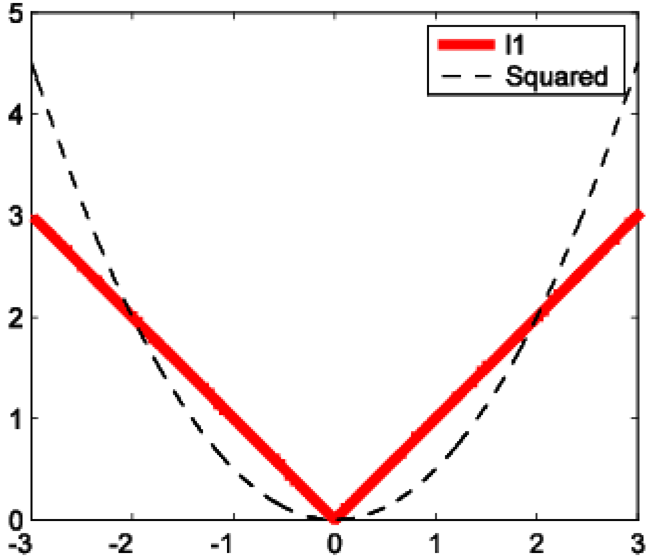
1. ℓ_1-損失
2. フーバー損失
3. テューキー損失

ℓ_1-損失を用いたロバスト回帰

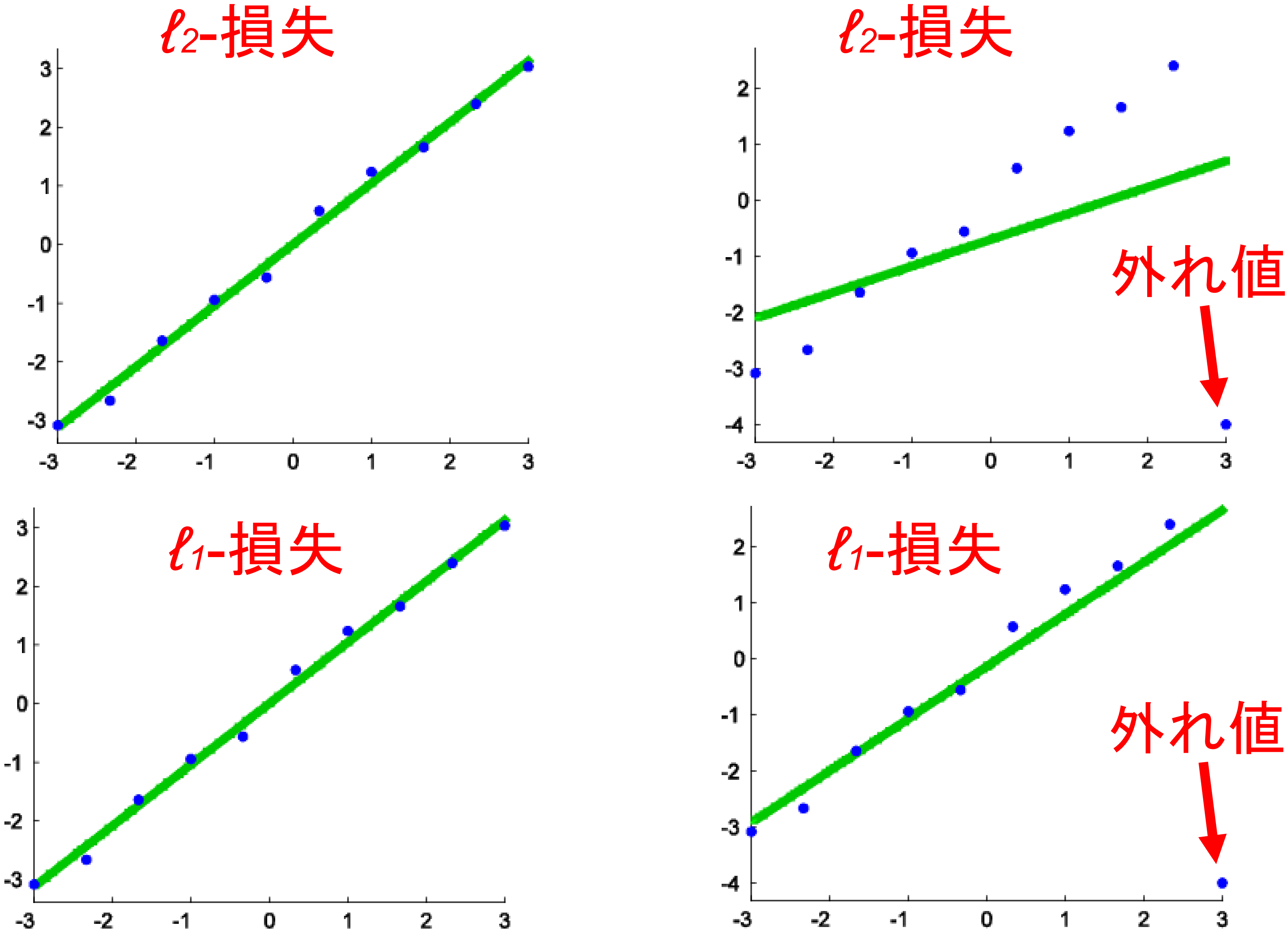
- ℓ_1-損失関数で訓練出力との適合のよさを測る:

$$\min_{\boldsymbol{\theta}} \sum_{i=1}^{n} \left| f_{\boldsymbol{\theta}}(\boldsymbol{x}_i) - y_i \right|$$

- 外れ値の影響は線形にしか効いてこない.



例



解の求め方は後ほど

累積分布関数

- 連続型の確率変数が x 以下の値をとる確率

$$P(x) = \mathrm{Prob}(X \leq x) = \int_{-\infty}^{x} p(u)du$$

$P(x)$:累積分布関数(cumulative distribution function)

$$P'(x) = \frac{\mathrm{d}P(x)}{\mathrm{d}x} = p(x)$$

- 広義単調増加:

$$x_1 < x_2 \quad \Longrightarrow \quad P(x_1) \leq P(x_2)$$

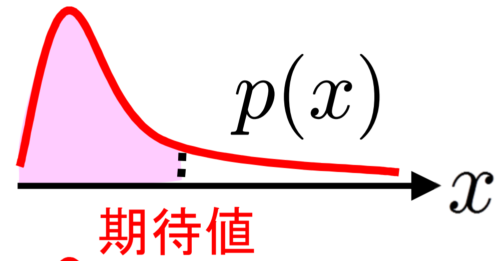
- 範囲:

$$x \to -\infty \quad \Longrightarrow \quad P(x) \to 0$$

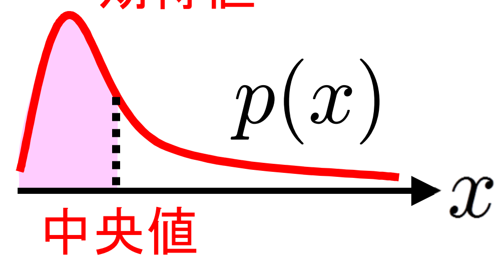
$$x \to \infty \quad \Longrightarrow \quad P(x) \to 1$$

期待値と中央値

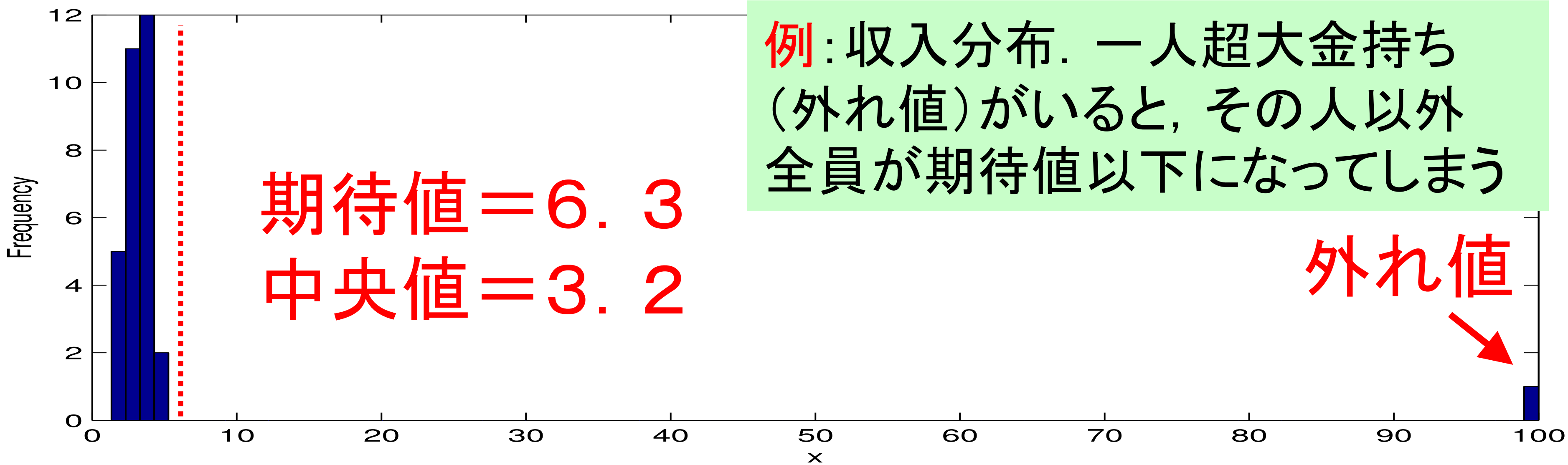
- **期待値**: $E[X] = \int xp(x)dx$



- **中央値**: $\mathrm{Prob}(X \leq x) = \frac{1}{2}$ を満たすx



- 期待値は，外れ値（outlier）があるときに直感と合わない値になることがある



例：収入分布．一人超大金持ち（外れ値）がいると，その人以外全員が期待値以下になってしまう

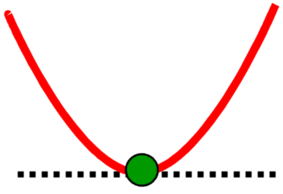
数学演習

- $[a,b]$上に定義された確率密度関数 $p(x)$ を考える
- 次の二乗誤差 $J_2(y)$ を最小にする y_2 は, x の期待値であることを示せ

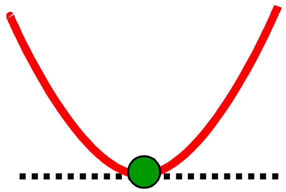
$$y_2 = \underset{y}{\operatorname{argmin}}\, J_2(y)$$

$$J_2(y) = \int_a^b (x-y)^2 p(x)dx$$

- ヒント: 最小点での微分はゼロ



解答例



- 最小点では微分がゼロなので,

$$J_2'(y_2) = 2\int_a^b (-x + y_2)p(x)dx = 0$$

$$\Longrightarrow \quad y_2 = \int_a^b xp(x)dx$$

数学演習

- $[a,b]$ 上の確率密度関数 $p(x)$ を考える
- 絶対値誤差 $J_1(y)$ を最小にする y_1 は x の中央値である($P(y_1)=1/2$ となる)ことを示せ

$$y_1 = \operatorname*{argmin}_y J_1(y)$$

$$J_1(y) = \int_a^b |x-y| p(x) dx$$

- ヒント: 累積分布の微分 $P'(x)=p(x)$ について部分積分を適用

$$\int_a^b f(x)g'(x)dx = \Big[f(x)g(x)\Big]_a^b - \int_a^b f'(x)g(x)dx$$

解答例

- 累積分布関数 $P(x)$ は $P(a) = 0,\ P(b) = 1$ を満たす
- 絶対値を展開する:

$$J_1(y) = \int_a^y (y-x)p(x)dx + \int_y^b (x-y)p(x)dx$$

- 部分積分を使うと

$$P'(x) = p(x)$$

$$J_1(y) = \Big[(y-x)P(x)\Big]_a^y + \int_a^y P(x)dx$$

$$+\Big[(x-y)P(x)\Big]_y^b - \int_y^b P(x)dx$$

解答例

- $Q'(x) = P(x)$ とおくと，　$P(a) = 0,\ P(b) = 1$

$$J_1(y) = \Big[(y-x)P(x)\Big]_a^y + \int_a^y P(x)dx$$

$$+\Big[(x-y)P(x)\Big]_y^b - \int_y^b P(x)dx$$

$$= Q(y) - Q(a) + b - y - Q(b) + Q(y)$$

- 最小点では微分がゼロなので，

$$J_1'(y_1) = 2P(y_1) - 1 = 0 \quad \Rightarrow \quad y_1 : P(y_1) = \frac{1}{2}$$

ロバスト学習の統計的な解釈

- 観測値＝真値＋雑音: $\{y_i \mid y_i = \mu^* + \epsilon_i\}_{i=1}^n$
- ℓ_2-損失関数: 観測値の期待値の推定に対応

$$\underset{\mu}{\operatorname{argmin}} \left[\sum_{i=1}^n (y_i - \mu)^2 \right] = \operatorname{mean}\left(\{y_i\}_{i=1}^n\right)$$

$$\to \mu^* + \operatorname{mean}(\epsilon)$$

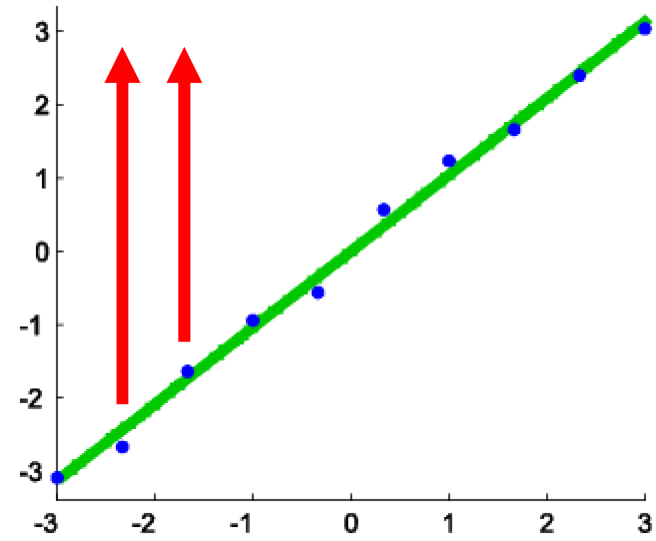
- ℓ_1-損失関数: 観測値の中央値の推定に対応

$$\underset{\mu}{\operatorname{argmin}} \left[\sum_{i=1}^n |y_i - \mu| \right] = \operatorname{median}\left(\{y_i\}_{i=1}^n\right)$$

$$\to \mu^* + \operatorname{median}(\epsilon)$$

ロバスト性と有効性

- 破綻点：標本のいくつかを無限に飛ばしたとき，学習結果が有限に留まる(破綻しない)標本数の割合の上限
  - ℓ_2-損失関数：0%
    ➡ 全くロバストでない
  - ℓ_1-損失関数：50%
    ➡ 非常にロバスト

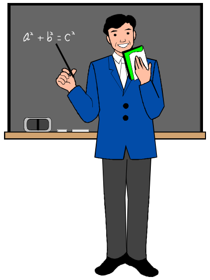


- ただし，ℓ_1-損失関数はガウス雑音に対して有効性を満たさない(分散が大きい)

ロバスト回帰における要請

- ロバスト性が高い ＝ 訓練標本をきちんと見ない
  - 常に0を出力する無意味な学習が最もロバスト

- 現実的な要請：
  - 外れ値がない場合には最小二乗法に近い
  - 外れ値が多い or 大きい場合にロバスト



講義の流れ

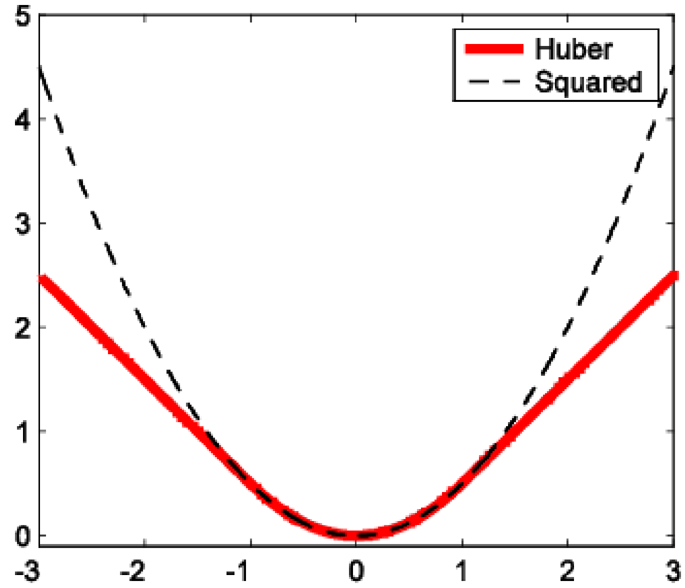
1. ℓ_1-損失
2. フーバー損失
3. テューキー損失

フーバー損失

■ ℓ_1-損失とℓ_2-損失の折衷案:フーバー損失

- 小さな誤差に対しては二乗
- 大きな誤差に対しては絶対値

$$\min_{\boldsymbol{\theta}} \sum_{i=1}^{n} \rho(f_{\boldsymbol{\theta}}(\boldsymbol{x}_i) - y_i)$$



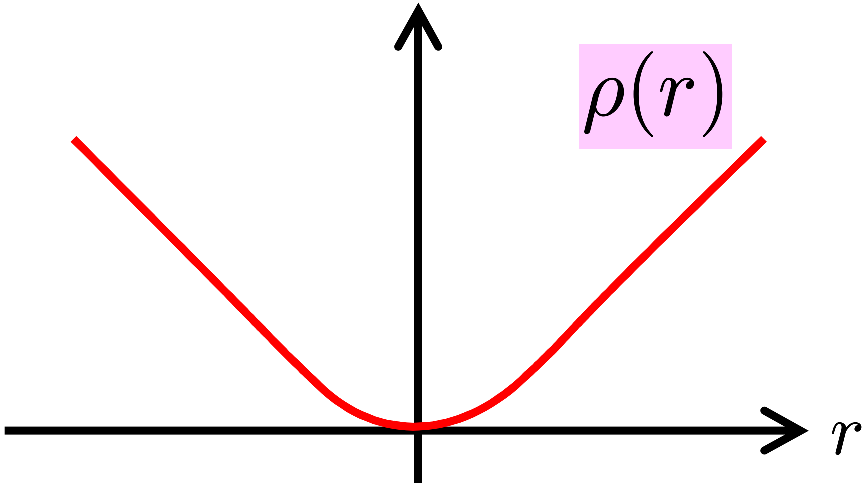
r

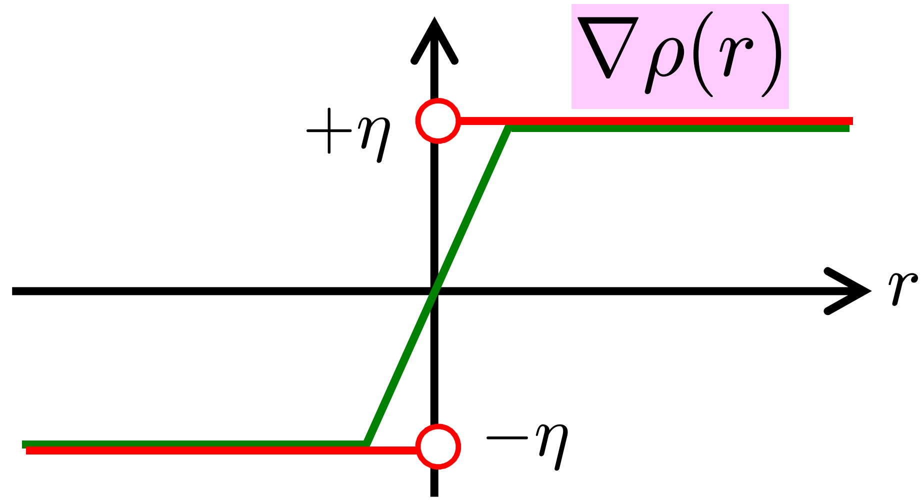
$$\rho(r) = \begin{cases} r^2/2 & (|r| \leq \eta) \\ \eta|r| - \eta^2/2 & (|r| > \eta) \end{cases}$$

$$\eta \geq 0$$

解の求め方1

- フーバー損失は連続的微分可能





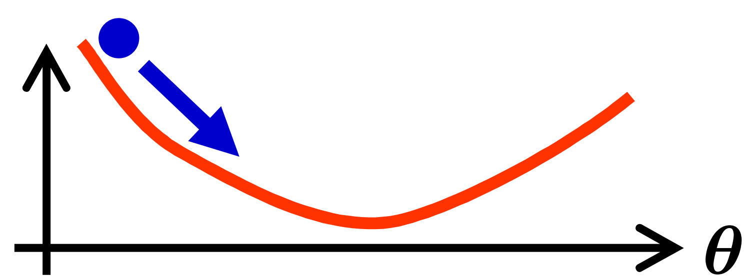
$$\rho(r) = \begin{cases} r^2/2 & (|r| \leq \eta) \\ \eta|r| - \eta^2/2 & (|r| > \eta) \end{cases}$$

$$\rho'(r) = \begin{cases} r & (|r| \leq \eta) \\ \mathrm{sign}(r)\eta & (|r| > \eta) \end{cases}$$

- 勾配法:

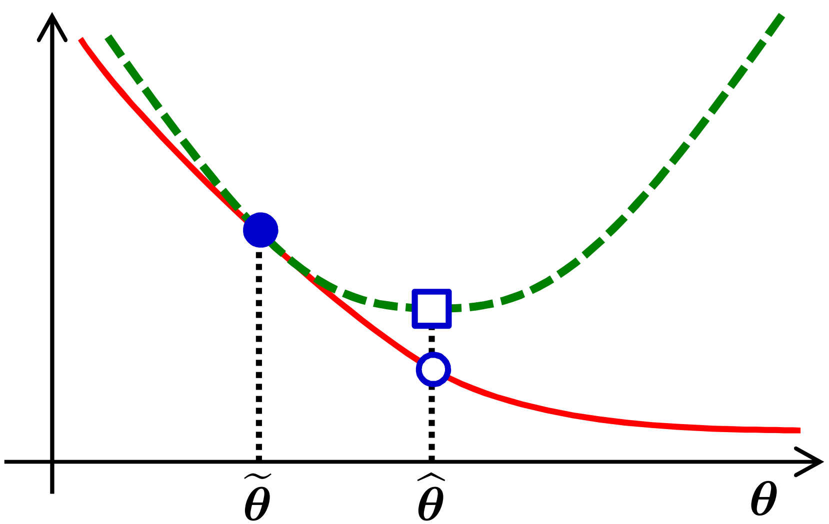
$$\boldsymbol{\theta} \longleftarrow \boldsymbol{\theta} - \varepsilon \nabla J(\boldsymbol{\theta})$$

$$J(\boldsymbol{\theta}) = \sum_{i=1}^{n} \rho(f_{\boldsymbol{\theta}}(\boldsymbol{x}_i) - y_i)$$



解の求め方2

- 勾配法はステップ幅の調整が厄介
- 繰り返し最小二乗アルゴリズム:
  - フーバー損失を現在の解で接するような二次関数で上から抑える
  - 二次上界を解析的に最小化することにより, 少しずつ良い解を求めていく



数学演習

- $|r| > \eta$ の場合のフーバー損失:

$$\eta|r| - \frac{\eta^2}{2}$$

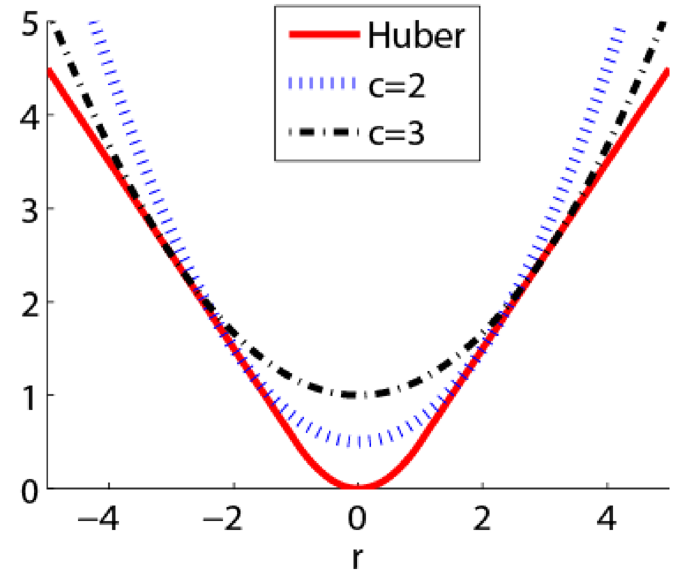
$$\rho(r) = \begin{cases} r^2/2 & (|r| \leq \eta) \\ \eta|r| - \eta^2/2 & (|r| > \eta) \end{cases}$$

に $r = \pm c\ , c > \eta$ で接する二次関数を求めよ

- ヒント: $\pm c$ で接する二次関数は対称性より

$$ar^2 + b$$

と表される



解答例

$$\eta|r| - \frac{\eta^2}{2} \le ar^2 + b$$

- $r = \pm c\ , c > \eta$ で傾きが等しい:

$$\pm\eta = \pm 2ac \Rightarrow a = \frac{\eta}{2c}$$

- $r = \pm c\ , c > \eta$ で接する:

$$\eta c - \frac{\eta^2}{2} = ac^2 + b \Rightarrow b = \eta c - \frac{\eta^2}{2} - ac^2$$

$$= \frac{\eta c}{2} - \frac{\eta^2}{2}$$

- 二次上界: $\eta|r| - \dfrac{\eta^2}{2} \le \dfrac{\eta}{2c} r^2 + \dfrac{\eta c}{2} - \dfrac{\eta^2}{2} \quad \text{for } c > \eta$

二次上界の最小化

$$\rho(r) = \begin{cases} r^2/2 & (|r| \le \eta) \\ \eta|r| - \eta^2/2 & (|r| > \eta) \end{cases}$$

- 現在の解 $\widetilde{\theta}$ に対する残差 $\widetilde{r} = f_{\widetilde{\boldsymbol{\theta}}}(\boldsymbol{x}) - y$ に対する二次上界 $\widetilde{\rho}(r) \ge \rho(r)$:

$$\widetilde{\rho}(r) = \begin{cases} r^2/2 & (|\widetilde{r}| \le \eta) \\ \dfrac{\eta}{2|\widetilde{r}|} r^2 + \underbrace{\dfrac{\eta|\widetilde{r}|}{2} - \dfrac{\eta^2}{2}}_{\text{定数}} & (|\widetilde{r}| > \eta) \end{cases}$$

$$= \frac{\widetilde{w}}{2} r^2 + \mathrm{const}$$

$$\widetilde{w} = \begin{cases} 1 & (|\widetilde{r}| \le \eta) \\ \eta/|\widetilde{r}| & (|\widetilde{r}| > \eta) \end{cases}$$

二次上界の最小化(続き)

- 最小化したいもとの損失:

$$J(\boldsymbol{\theta}) = \sum_{i=1}^{n} \rho(f_{\boldsymbol{\theta}}(\boldsymbol{x}_i) - y_i)$$

$$\rho(r) = \begin{cases} r^2/2 & (|r| \leq \eta) \\ \eta|r| - \eta^2/2 & (|r| > \eta) \end{cases}$$

- 現在の解 $\widetilde{\boldsymbol{\theta}}$ から求めた J の上界 $\widetilde{J}$ の最小化:

$$\widehat{\boldsymbol{\theta}} = \mathop{\mathrm{argmin}}_{\boldsymbol{\theta}} \widetilde{J}(\boldsymbol{\theta})$$

$$\widetilde{J}(\boldsymbol{\theta}) = \frac{1}{2} \sum_{i=1}^{n} \widetilde{w}_i \Big(f_{\boldsymbol{\theta}}(\boldsymbol{x}_i) - y_i \Big)^2$$

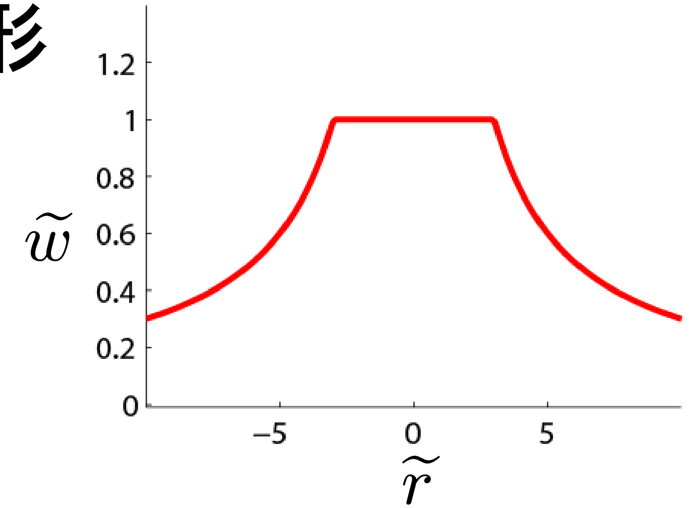
$$\widetilde{w}_i = \begin{cases} 1 & (|\widetilde{r}_i| \leq \eta) \\ \eta/|\widetilde{r}_i| & (|\widetilde{r}_i| > \eta) \end{cases}$$

$$\widetilde{r}_i = f_{\widetilde{\boldsymbol{\theta}}}(\boldsymbol{x}_i) - y_i$$

二次上界の最小化(続き)

- 上界は重み付き最小二乗法の形

$$\min_{\boldsymbol{\theta}} \frac{1}{2}\sum_{i=1}^{n}\widetilde{w}_i\Big(f_{\boldsymbol{\theta}}(\boldsymbol{x}_i)-y_i\Big)^2$$



  - 異常値に対する重みを小さく設定

$$\widetilde{w}_i=\begin{cases}1 & (|\widetilde{r}_i|\le\eta)\\ \eta/|\widetilde{r}_i| & (|\widetilde{r}_i|>\eta)\end{cases}$$

$$\widetilde{r}_i=f_{\widetilde{\boldsymbol{\theta}}}(\boldsymbol{x}_i)-y_i$$

- 上界の最小解は解析的に次式で求められる

$$\widehat{\boldsymbol{\theta}}=(\boldsymbol{\Phi}^\top\widetilde{\boldsymbol{W}}\boldsymbol{\Phi})^{-1}\boldsymbol{\Phi}^\top\widetilde{\boldsymbol{W}}\boldsymbol{y}$$

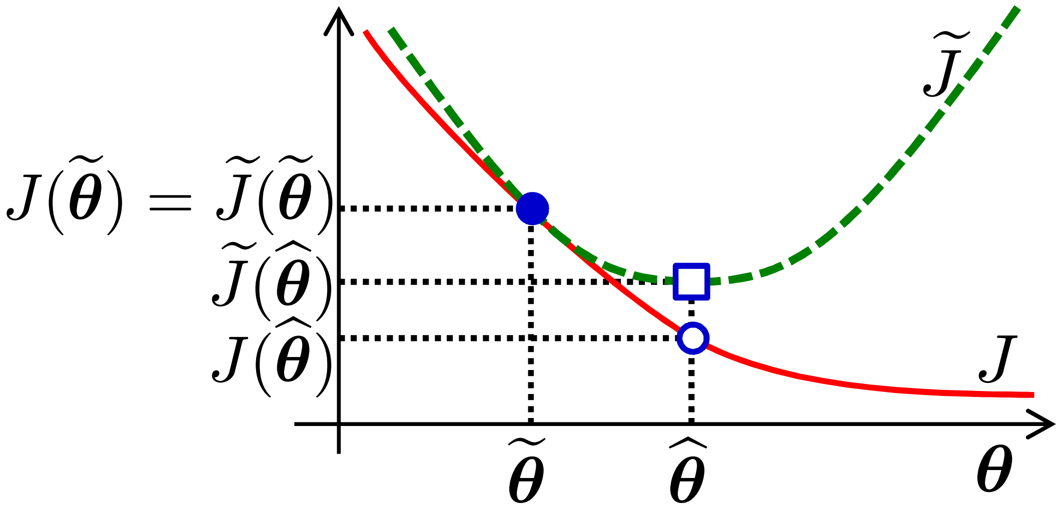
$$\widetilde{\boldsymbol{W}}=\mathrm{diag}\left(\widetilde{w}_1,\ldots,\widetilde{w}_n\right)$$

証明は
宿題

二次上界の最小化(続き)

- 上界が $\widetilde{\boldsymbol{\theta}}$ で接することから $J(\widetilde{\boldsymbol{\theta}}) = \widetilde{J}(\widetilde{\boldsymbol{\theta}})$
- $\widehat{\boldsymbol{\theta}}$ を上界の最小解とすれば $\widetilde{J}(\widetilde{\boldsymbol{\theta}}) \geq \widetilde{J}(\widehat{\boldsymbol{\theta}})$
- 上界であることから $\widetilde{J}(\widehat{\boldsymbol{\theta}}) \geq J(\widehat{\boldsymbol{\theta}})$
- まとめると $J(\widetilde{\boldsymbol{\theta}}) = \widetilde{J}(\widetilde{\boldsymbol{\theta}}) \geq \widetilde{J}(\widehat{\boldsymbol{\theta}}) \geq J(\widehat{\boldsymbol{\theta}})$

解を $\widetilde{\boldsymbol{\theta}}$ から $\widehat{\boldsymbol{\theta}}$ に更新すれば J の値は(一般に)減少する



$$\widehat{\boldsymbol{\theta}} = \operatorname*{argmin}_{\boldsymbol{\theta}} \widetilde{J}(\boldsymbol{\theta})$$

繰り返し再重み付け最小二乗アルゴリズム

- $\boldsymbol{\theta}$ を適当に初期化する
- 以下を収束するまで繰り返す
  - 現在の解 $\boldsymbol{\theta}$ から行列 $\boldsymbol{W}$ を計算する（上界を求める）

$$\boldsymbol{W} = \mathrm{diag}\left(w_1, \ldots, w_n\right)$$

$$w_i = \begin{cases} 1 & (|f_{\boldsymbol{\theta}}(\boldsymbol{x}_i) - y_i| \leq \eta) \\ \eta / |f_{\boldsymbol{\theta}}(\boldsymbol{x}_i) - y_i| & (|f_{\boldsymbol{\theta}}(\boldsymbol{x}_i) - y_i| > \eta) \end{cases}$$

  - 解 $\boldsymbol{\theta}$ を更新する（上界を最小化する）

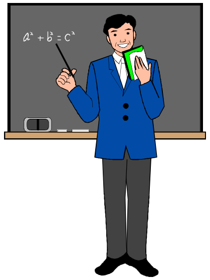
$$\boldsymbol{\theta} \longleftarrow (\boldsymbol{\Phi}^\top \boldsymbol{W} \boldsymbol{\Phi})^{-1} \boldsymbol{\Phi}^\top \boldsymbol{W} \boldsymbol{y}$$

実装例

- 直線モデル $f_{\boldsymbol{\theta}}(x) = \theta_1 + \theta_2 x$ に対するフーバー回帰の繰り返し最小二乗アルゴリズム

```
clear all; rand('state',0); randn('state',0);
n=10; N=1000;
x=linspace(-3,3,n)'; X=linspace(-4,4,N)';
y=x+0.2*randn(n,1); y(n)=-4;
p(:,1)=ones(n,1); p(:,2)=x; t0=p¥y; e=1;
for o=1:1000
  r=abs(p*t0-y); w=ones(n,1); w(r>e)=e./r(r>e);
  t=(p'*(repmat(w,1,2).*p))¥(p'*(w.*y));
  if norm(t-t0)<0.001, break, end
  t0=t;
end
P(:,1)=ones(N,1); P(:,2)=X; F=P*t;
figure(1); clf; hold on; axis([-4 4 -4.5 3.5]);
plot(X,F,'g-'); plot(x,y,'bo');
```

η が小さい場合
ℓ_1-損失最小化と
ほぼ一致

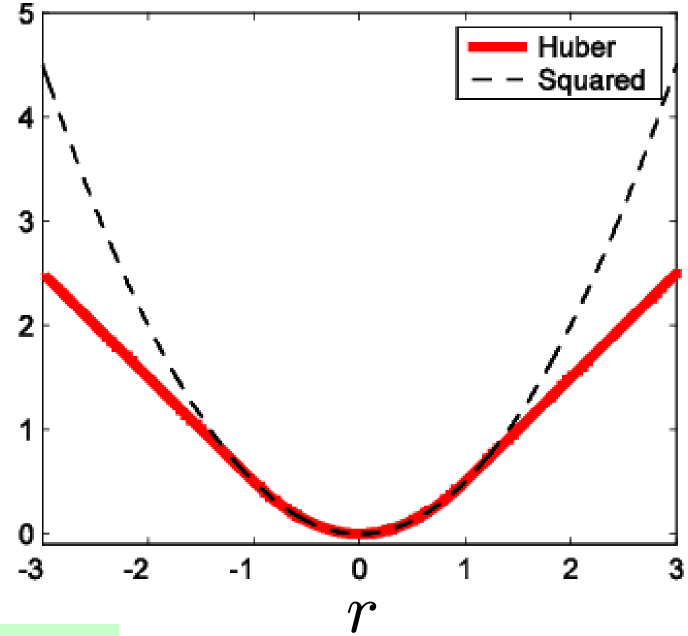


講義の流れ

1. ℓ_1-損失
2. フーバー損失
3. テューキー損失

強い外れ値の影響

- フーバー損失はℓ_2-損失と比べてロバスト性が高い
- しかし，損失に上界がないため
  非常に強い外れ値の影響は
  受けてしまう



$$\min_{\boldsymbol{\theta}} \sum_{i=1}^{n} \rho_{\mathrm{Huber}}(f_{\boldsymbol{\theta}}(\boldsymbol{x}_i) - y_i)$$

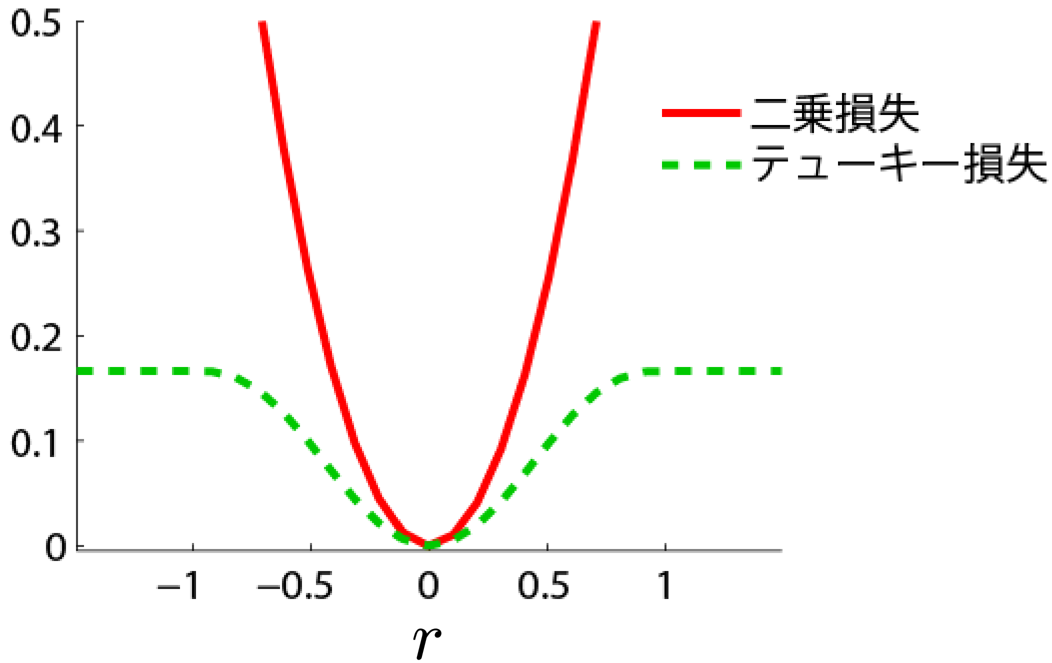
$$\rho_{\mathrm{Huber}}(r) = \begin{cases} r^2/2 & (|r| \leq \eta) \\ \eta|r| - \eta^2/2 & (|r| > \eta) \end{cases}$$

テューキー損失

- 上界のある損失を考える

$$r = f_{\boldsymbol{\theta}}(\boldsymbol{x}) - y$$

$$\rho_{\mathrm{Tukey}}(r) = \begin{cases} \left(1 - \left[1 - r^2/\eta^2\right]^3\right)/6 & (|r| \leq \eta) \\ 1/6 & (|r| > \eta) \end{cases}$$



一般の損失に対する二次上界の最小化

$$\min_{\boldsymbol{\theta}} \sum_{i=1}^{n} \rho(f_{\boldsymbol{\theta}}(\boldsymbol{x}_i) - y_i)$$

- 微分可能で対称な損失 $\rho(r)$ に対して $\widetilde{r}$ で接する二次上界は(存在するなら)次式で与えられる:

$$\widetilde{\rho}(r) = \frac{\widetilde{w}}{2} r^2 + \text{const}$$

$$\widetilde{w} = \rho'(\widetilde{r})/\widetilde{r}$$

証明は宿題

- 繰り返し最小二乗アルゴリズム:

$$\min_{\boldsymbol{\theta}} \frac{1}{2} \sum_{i=1}^{n} \widetilde{w}_i \Big(f_{\boldsymbol{\theta}}(\boldsymbol{x}_i) - y_i \Big)^2$$

$$\widetilde{w}_i = \rho'(\widetilde{r}_i)/\widetilde{r}_i$$

$$\widetilde{r}_i = f_{\widetilde{\boldsymbol{\theta}}}(\boldsymbol{x}_i) - y_i$$

テューキー損失に対する重み

■テューキー損失

$$\rho_{\mathrm{Tukey}}(r) = \begin{cases} \left(1 - \left[1 - r^2/\eta^2\right]^3\right)/6 & (|r| \le \eta) \\ 1/6 & (|r| > \eta) \end{cases}$$

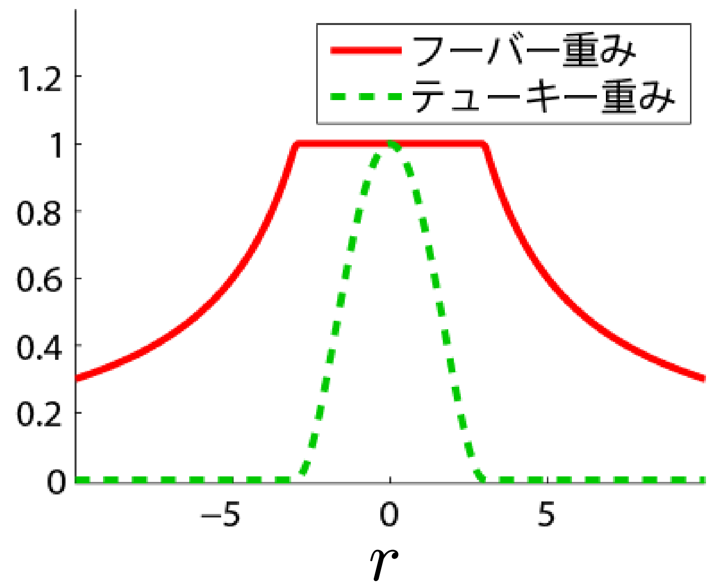
に対する重みは次式で与えられる:

$$\widetilde{w} = \rho'(\widetilde{r})/\widetilde{r}$$

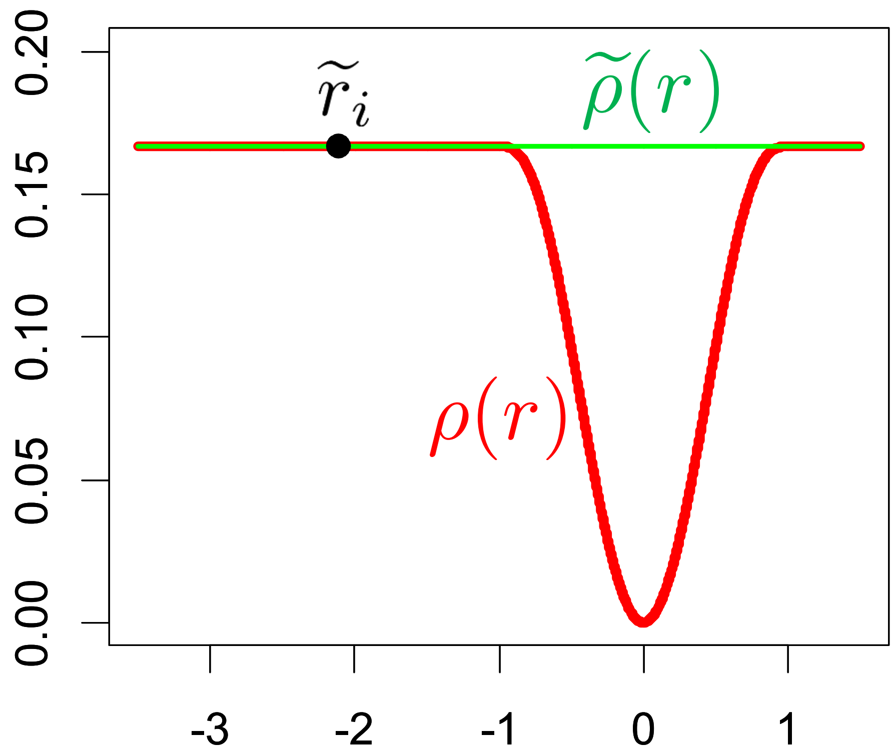
$$w_{\mathrm{Tukey}} = \begin{cases} \left(1 - r^2/\eta^2\right)^2 & (|r| \le \eta) \\ 0 & (|r| > \eta) \end{cases}$$

- 大きな残差に対する重みがゼロ

$$w_{\mathrm{Huber}} = \begin{cases} 1 & (|r| \le \eta) \\ \eta/|r| & (|r| > \eta) \end{cases}$$

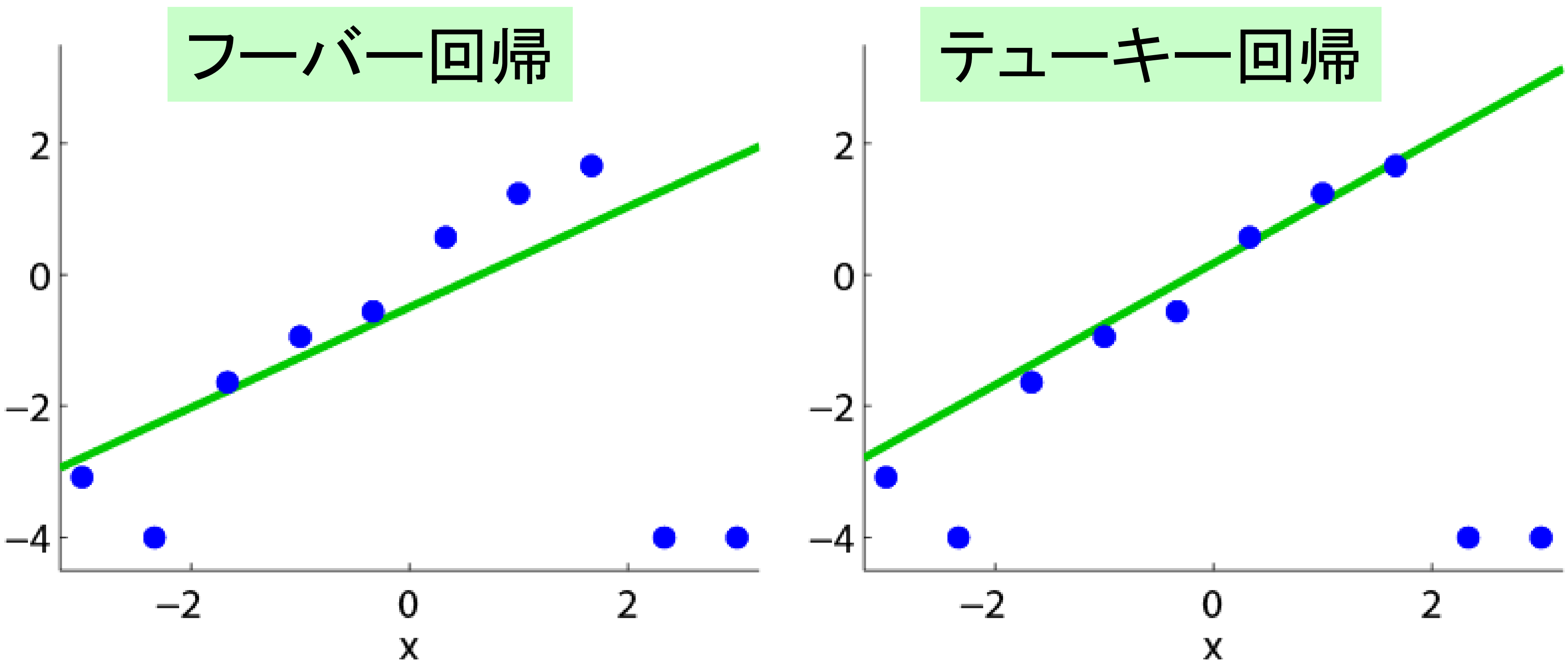


損失関数の非凸性



- 非凸な損失関数の二次上界をとった場合は最適解方向に進めるとは限らない

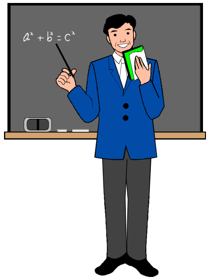
実行例



- テューキー回帰の方が外れ値に強い
- ただし，非凸最適化のため，得られる解は初期値の選び方に依存する

ロバスト回帰：まとめ

- 二乗損失（平均値）は外れ値に弱い
- 絶対値損失（中央値）は外れ値に強い
- フーバー損失はロバスト性と有効性のバランスがとれている
  - 解は解析的に求められない
- テューキー損失を用いるとロバスト性が更に向上
  - ただし非凸最適化になり最適化が困難



回帰のまとめ

1. 学習モデル(2章)
2. 最小二乗回帰(3章)
3. 正則化回帰(4章)
4. スパース回帰(5章)
5. ロバスト回帰(6章)

回帰のまとめ

- 関数を学習するためのモデル:
  - 線形モデル
  - カーネルモデル
  - 非線形モデル
- 最小二乗回帰:
  - 訓練標本との二乗誤差を最小化
  - 解は解析的に計算できる
- オンライン回帰
  - 多数の標本を逐次的に学習
  - 実装が非常に簡単

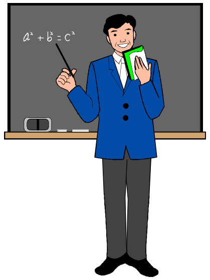
回帰のまとめ(続き)

- ℓ_2-正則化回帰:
  - 最小二乗回帰の過適合を軽減
  - 解は解析的に計算できる
  - モデル選択には交差確認法を用いる
- ℓ_1-正則化回帰(スパース回帰)
  - 真のパラメータ値の多くがゼロとなるデータを適切に学習
- ロバスト回帰
  - 異常な値に対するロバスト性を強化

回帰のまとめ（続き）

■ 線形モデル／カーネルモデルに対する回帰法：

損失関数 ＼ 制約条件	無し	ℓ_2-制約 正則化	ℓ_1-制約 正則化＆スパース
ℓ_2-損失　有効	解析解	解析解	二次計画
フーバー損失	二次計画	二次計画	二次計画
ℓ_1-損失　ロバスト	線形計画	二次計画	線形計画



次回の予告

- 最小二乗分類(7章)

宿題1

- 線形モデル

$$f_{\boldsymbol{\theta}}(\boldsymbol{x}) = \sum_{j=1}^{b} \theta_j \phi_j(\boldsymbol{x})$$

$\{\phi_j(\boldsymbol{x})\}_{j=1}^{b}$
:基底関数

に対する重み付き最小二乗法

$$\min_{\boldsymbol{\theta}} \frac{1}{2} \sum_{i=1}^{n} \widetilde{w}_i \Big(f_{\boldsymbol{\theta}}(\boldsymbol{x}_i) - y_i \Big)^2$$

の解が次式で与えられることを示せ:

$$\widehat{\boldsymbol{\theta}} = (\boldsymbol{\Phi}^\top \widetilde{\boldsymbol{W}} \boldsymbol{\Phi})^{-1} \boldsymbol{\Phi}^\top \widetilde{\boldsymbol{W}} \boldsymbol{y}$$

$$\boldsymbol{\Phi} = \begin{pmatrix} \phi_1(\boldsymbol{x}_1) & \cdots & \phi_b(\boldsymbol{x}_1) \\ \vdots & \ddots & \vdots \\ \phi_1(\boldsymbol{x}_n) & \cdots & \phi_b(\boldsymbol{x}_n) \end{pmatrix}$$

$$\widetilde{\boldsymbol{W}} = \mathrm{diag}\left(\widetilde{w}_1, \ldots, \widetilde{w}_n\right)$$

$$\boldsymbol{y} = (y_1, \ldots, y_n)^\top$$

宿題2

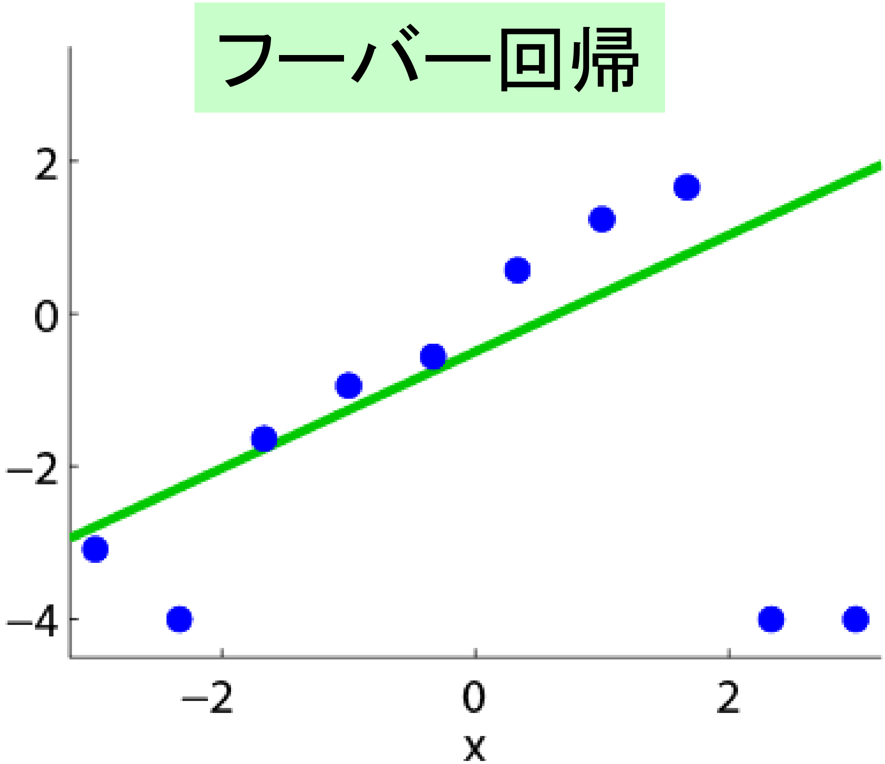
- 微分可能で対称な損失 $\rho(r)$ に対して $\widetilde{r}$ で接する二次上界は（存在するなら）次式で与えられることを示せ

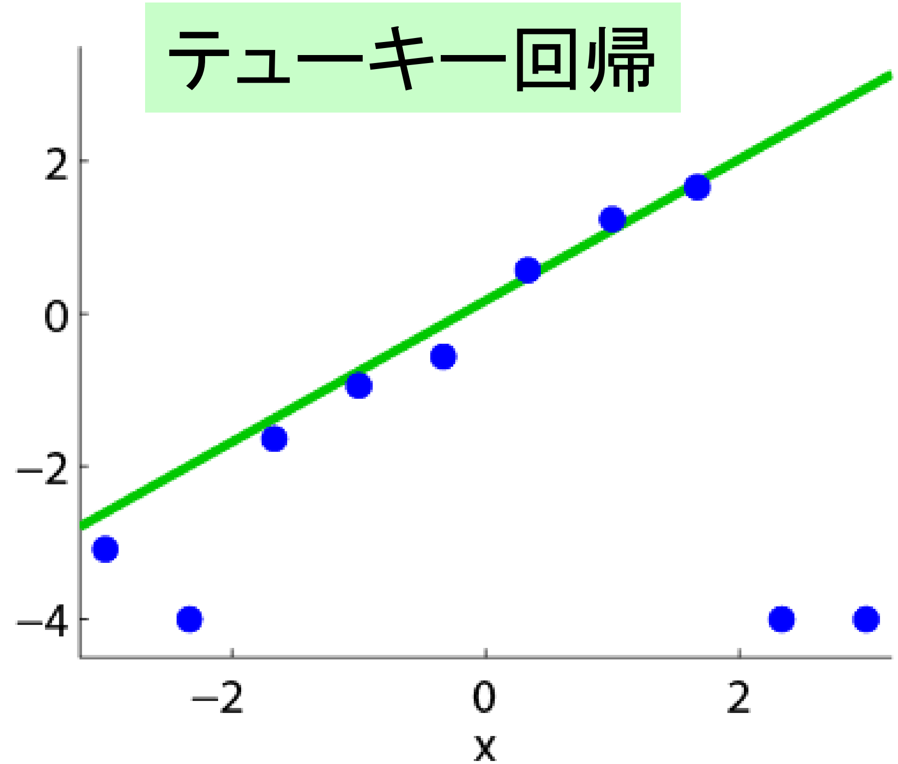
$$\widetilde{\rho}(r) = \frac{\widetilde{w}}{2} r^2 + \mathrm{const} \qquad \widetilde{w} = \rho'(\widetilde{r})/\widetilde{r}$$

- const : r に依存しない値

宿題3

- 直線モデル $f_{\boldsymbol{\theta}}(x) = \theta_1 + \theta_2 x$ に対して，テューキー回帰の繰り返し最小二乗アルゴリズムを実装せよ（結果は初期値に依存して変わる場合があることに注意）





宿題3の続き

- データとしては例えば以下のものを用いてもよい

```
clear all; rand('state',0); randn('state',0);
n=10; N=1000; x=linspace(-3,3,n)';
y=x+0.2*randn(n,1); y(n)=-4; y(n-1)=-4; y(2)=-4;
```